
NORTH CAROLINA

**EDUCATIONAL COMPUTING SERVICE**

# THE POISSON PRIMER

## PACKAGE OF INSTRUCTIONAL SOCIAL SURVEYS OF NORTH CAROLINA

J. R. DENK AND N. MOXLEY, NCECS



APRIL 1, 1972

Established by the **North Carolina Board of Higher Education**

# The POISSON Primer

Package of Instructional Social Surveys
of North Carolina

J. R. Denk and N. Moxley, NCECS

April 1, 1972

North Carolina Educational Computing Service
Post Office Box 12175
Research Triangle Park, North Carolina 27709

TABLE OF CONTENTS

Chapter One

# A POISSON SCENARIO*

## Overture

The following scenes are sketched to show how the POISSON system can be used in the classroom. At its exciting climax, we present a real POISSON run and explain the major steps involved. Other sample runs can be seen in Chapter Five.

## Act 1

Curtain rises on a UNC-Greensboro classroom, presided over by Professor Knocks, who is speaking to a group of 25 young Carolinians, trying to find political relevance in a blank biennium.

PK: How many of you had a chance to look at today's assignment, pages 219-249 of Robert Alford's Party and Society, Rand McNally & Company, Chicago, 1963? (Professor Knocks is a bit of a pedant.)

A forest of hands pop up (remember, this is fiction).

PK: Well, now what seems to be his main argument? Can you tell us, Mr. Striver?

ERNEST STRIVER: Well, Professor Knocks, Sir, Alford argues that blue collar people are more likely to vote for the more liberal party, that is the Democrats, and that this difference is essentially constant over time.

PK: Good work there, Striver, but doesn't he refine his argument further?

ES: As a matter of fact, he does just that, as you said Professor Knocks, Sir. He further argues that this difference is not as strong among Catholics and not as strong in the South (the hushed silence now becomes deafening). He attributes such a specification--as we call it in technical lingo--to subcultural differences.

---

*This scenario was adapted from the PRIMER of Project IMPRESS at Dartmouth College. Hopefully, enough ivy has been cut to produce a Tarheel classroom atmosphere. Other chapters in the primer have also borrowed heavily from the Project IMPRESS primer.

(PK totters over to the blackboard and chalks up the following figures.)

Difference (Per Cent Democratic among Blue Collar Minus Per Cent Democratic among White Collar)

| Year | | Total | Religion | | | Region | |
|---|---|---|---|---|---|---|---|
| | | | Cath. | Prot. | P-C | South | Total-South |
| 1936 | | +16 | | | | | |
| 1940 | | +25 | | | | | |
| 1944 | | +20 | +19 | +23 | +4 | +11 | +9 |
| 1948 | | +20 | +8 | +25 | +17 | +16 | +4 |
| 1952 | sample 1 | +22 | +26 | +24 | -2 | +10 | +12 |
| | sample 2 | +16 | +11 | +20 | +9 | +11 | +6 |
| 1956 | | +16 | +11 | +20 | +9 | +3 | +13 |
| 1960 | sample 1 | +14 | +6 | +19 | +13 | +9 | +5 |
| | sample 2 | +15 | +10 | +18 | +8 | +18 | -3 |

After a period of discussion, not unlike extrative dentistry, the class comes to agree with Striver's interpretation. The column headed "TOTAL" does show a fairly constant occupational difference. The column headed "P-C" shows this difference to be greater among Protestants in all but one case, and the column headed "Total-South" shows the difference to be greater in the total population among Southerners (these data, by the way, are for real; but Alford's published tables don't make it easy to get two separate figures for South and non-South; thus, we compare with the total U.S.).

PK: Well, now is this still true or have there been changes since 1960? After all, that's 12 years ago.

(The silence now is pierced by sotto voce grumbling that this wasn't in the reading assignment. After shock subsides, various tentative ideas are voiced.)

MR. X: I don't think so.

MR. Y: Well, there was this Playboy interview...

MISS Z: I have some Black friends at A and T who say . . .

(Professor Knocks waves them all aside.)

PK: Doesn't anybody have any hard evidence on this gripping question? (A hand goes up in the back of the room. Attached to it is Billie Joe Williams, a shy lass who had spoken little during the term so far. From her unfashionable knee length dress and short-clipped hair, Professor Knocks judges her to be a scholarship student from rural North Carolina.)

BJW: Ah think POISSON could give us the answer.

(All male heads are seen to flush jealously.)

Mr. X: Gollie! I should have thought of that myself.

Mr. Y: Wooee!!

PK: Tell us more, Miss Williams!

BJW: Well, there is a survey called PRES68 available through the POISSON system. It is a national sampling of American adults in Fall, 1968, and I'm sure it has questions about voting, region, occupation, and religion.

PK: I think you have something there, Miss Williams. I see we are getting to the end of the hour. Will you please run PRES68 for tomorrow and try to find out whether Alford's conclusions are still valid.

(EXIT all in a dither.)

Act 2

The Reserve Reading Room of the library. Billie Joe is discovered pouring over a PRES68 codebook. By subtle pantomime it is revealed that she finds region to be measured by a POISSON variable called REGIØN, religious preference by RELIG, occupation by ØCC1, and voting preference by VØTED. She jots some notes and exits running.

Act 3

The terminal room in the Science Hall: Billie Joe is undaunted by chemistry odors and the other foreign atmosphere. She sits at a keyboard terminal

(in this case a teletype), follows the usual sign-on procedures, and generates typed output.

Here, we depart from our drama to follow the actual run of POISSON. Billie Joe's input follows each question mark (?).

A
```
SERVICE?
CPS
?xxx.yyy.zzz,WILLIAMS
GØØD AFTERNØØN:  USER 09;  TIME  16:30:50  10/28/71
```

B
```
?LØAD(PØISØN)PUBLIC
?XEQ
DØ YØU WANT USER'S INSTRUCTIØNS?
?NØ
```

C
```
ENTER NAME ØF SURVEY
?PRES68
      1968 ELECTIØN SURVEY -- MICHIGAN SRC
ENTER NAMES ØF 2 TØ 8 VARIABLES TØ BE RUN
(COMMAS) - NØ BLANKS PLEASE.
?REGIØN,RELIG,ØCC1,VØTED
```

D
```
ENTER GRØUPING ØPTIØN
?SD,EXCEPT,RELIG,MAP,VØTED,MAP
```

A. Billie Joe follows the usual sign-on procedure for CPS (conversational programming system) in order to use the interactive version of POISSON.

B. She asks for the CPS library program, POISON. A question mark follows and she requests that POISON be executed (run). Being an old hand, Billie Joe by-passes the user's instructions.

C. The program asks her to type in the name of the study she wants (PRES68) and the names of the variables she will be working with (REGIØN,RELIG, ØCC1,VØTED).

D. The program asks her whether she wants to work with pre-defined category groupings within her variables or whether she wants to make her own groupings. Billie Joe tells it that the standard dichotomy for REGIØN (South v. All Other) and OCCUPATION (White Collar v. Blue Collar) are fine but she wants to make her own category groupings for RELIG and VØTED.

E

```
'RELIG' LØCATED--RELIGIØUS PREFERENCE

HØW MANY NEW CATEGØRIES?
?3

GRØUP 1?
?PRØT,0-4

GRØUP 2?
?CATH,5

GRØUP 3?
?ØTHNA,6-13

ENTER EXCLUSIØN CØDE
?1
```

F

```
'VØTED' LØCATED--VØTED ØR WØULD HAVE VØTED FØR WHØM IN 68

HØW MANY NEW CATEGØRIES?
?3

GRØUP 1?
?HHH,0,1,14,15

GRØUP 2?
?RMN,2,3,16,17

GRØUP 3?
?ØTHNA,4-13,18-34

ENTER EXCLUSIØN CØDE
?1
```

G

```
ENTER CØMMAND
?MARG:RELIG,VØTED

YØUR CØMMAND, VARIABLES, AND ØPTIØNS HAVE BEEN
ACCEPTED.  WØULD YØU LIKE TØ ENTER ANØTHER
CØMMAND (YES ØR NØ)?
?YES
```

E. The MAP option allows Billie Joe to dichotomize RELIG into Protestant (PRØT) and Catholic (CATH) while lumping what responses are to be excluded into a third group, ØTHNA. She then enters an EXCLUSIØN CØDE of 1 which excludes from analysis the group with the highest number (group 3 in this case). (An exclusion code of 0 would exclude no groups, 1 would exclude the group with the highest number, 2 would exclude the two groups with the highest numbers, etc.)

F. She then goes through the same process for the variable, VØTED.

G. The program tells Billie Joe that it has her variables all arranged and asks for instructions with the request, ENTER CØMMAND. She first asks for the marginals (one-variable frequency distributions) for her home-made variables, RELIG and VØTED. Since the interactive version of POISSON is built to direct the user in how to punch cards to get the answer, no answer is given her at this point. The program asks her if she would like to pose other questions of the variables she has chosen.

```
ENTER CØMMAND
?XTAB:ØCC1,VØTED;%A

YØUR CØMMAND, VARIABLES, AND ØPTIØNS HAVE BEEN
ACCEPTED.  WØULD YØU LIKE TØ ENTER ANØTHER
CØMMAND (YES ØR NØ)?
?YES

ENTER CØMMAND
?XTAB:RELIG,ØCC1,VØTED;%A

YØUR CØMMAND, VARIABLES, AND ØPTIØNS HAVE BEEN
ACCEPTED.  WØULD YØU LIKE TØ ENTER ANØTHER
CØMMAND (YES ØR NØ)?
?YES

ENTER CØMMAND
?XTAB:REGIØN,ØCC1,VØTED;%A

YØUR CØMMAND, VARIABLES, AND ØPTIØNS HAVE BEEN
ACCEPTED.  WØULD YØU LIKE TØ ENTER ANØTHER
CØMMAND (YES ØR NØ)?
?NØ
```

H

```
      THE FØLLØWING IS A PRINT-ØUT ØF YØUR CØNTRØL CARDS FØR
A 'PØISSØN' RUN.  THE FIRST TWØ LINES ARE CARD CØLUMNS AND ARE
NØT TØ BE PUNCHED.

         1    1    2    2    3    3    4    4    5    5    6    6    7
....5....0....5....0....5....0....5....0....5....0....5....0....5....0

ID          PRES68     REGIØN,RELIG,ØCC1,VØTED
GRØUPING    SD         RELIG=MAP,VØTED=MAP
MAP         RELIG,1    PRØT,0-4;CATH,5;ØTHNA,6-13
MAP         VØTED,1    HHH,0,1,14,15;RMN,2,3,16,17;ØTHNA,4-13,18-34
CØMMAND     MARG:RELIG,VØTED
CØMMAND     XTAB:ØCC1,VØTED;%A
CØMMAND     XTAB:RELIG,ØCC1,VØTED;%A
CØMMAND     XTAB:REGIØN,ØCC1,VØTED;%A

?LØGØUT
TIME 13:58:30  TIME USED: CPU 00:00:01;  TERM 00:14:10;  PAGE 00:52:31;
```

I

H. Billie Joe then gives three consecutive CØMMANDS to pull out data which will give her three entries to Professor Knocks' table for the 1968 survey. She requests the percentage preferring Humphrey and Nixon among White and Blue Collar Workers, first without a control variable, then controlling for RELIG, and finally controlling for REGIØN. No further CØMMANDS are requested.

I. Billie Joe is directed to the exact format for card input to the POISSON system. She also finds out she has been on the terminal about 14 minutes to get her questions into this format. Now, the "batch" version of POISSON is used to obtain the data requested. She punches up her cards and submits her runs. Output from the "batch" run follows.

The batch output of a POISSON run first echoes back the input cards. These are not repeated here for simplicity.

---

J

```
YØUR CØMMAND: MARG:RELIG,VØTED

                MARGINAL FREQUENCIES
                -------- -----------

RELIG       =   RELIGIØUS PREFERENCE

PRØT              1167      69.8%
CATH               348      20.8%
ØTHNA              158       9.4% EXCLUDED

VØTED       =   VØTED ØR WØULD HAVE VØTED FØR WHØM IN 68

HHH                635      37.9%
RMN                613      36.6%
ØTHNA              425      25.4% EXCLUDED

SAMPLE SIZE = 1673
```

---

K

```
YØUR CØMMAND: XTAB:ØCC1,VØTED;%A

ØCC1 BY VØTED

DØWN:    RESPØNDENT'S MAIN ØCCUPATIØN - PØL BEH CØDE
ACRØSS:  VØTED ØR WØULD HAVE VØTED FØR WHØM IN 68

                    PERCENTAGES ACRØSS
                    ---------- ------

                HHH           RMN           TØTAL

WHITE          45.8%         54.1%          423
BLUE           60.6%         39.3%          407

TØTAL          53.1%         46.8%          830

EXECUSIØN ANALYSIS:
TABLE TØTAL     =     830
EXCLUDED        =     843
SAMPLE SIZE     =    1673
```

60.6
-45.8
14.8

J. The first output Billie Joe receives are a table of marginals by which she receives some idea of how large her cell entries are before she continues analysis.

K. The second CØMMAND produces the percentages preferring Humphrey and Nixon among White and Blue Collar Workers. Subtracting 45.8 from 60.6, Billie Joe has produced another entry to Professor Knocks' table on the blackboard under the column TØTAL. This differential is the same, she says, as that of 1960!

YØUR CØMMAND: XTAB:RELIG,ØCC1,VØTED;%A

RELIG BY ØCC1 BY VØTED

CØNTRØL: RELIGIØUS PREFERENCE
DØWN: RESPØNDENT'S MAIN ØCCUPATIØN - PØL BEH CØDE
ACRØSS: VØTED ØR WØULD HAVE VØTED FØR WHØM IN 68

PERCENTAGE ACRØSS

RELIG = PRØT

| | HHH | RMN | TØTAL |
|---|---|---|---|
| WHITE | 37.1% | 62.8% | 277 |
| BLUE | 61.4% | 38.5% | 306 |
| TØTAL | 49.9% | 50.0% | 583 |

61.4
-37.1
24.3

RELIG = CATH

| | HHH | RMN | TØTAL |
|---|---|---|---|
| WHITE | 61.9% | 38.0% | 105 |
| BLUE | 55.2% | 44.7% | 76 |
| TØTAL | 59.1% | 40.8% | 181 |

55.2
-61.9
-6.7

EXECUSIØN ANALYSIS:
TABLE TØTAL = 764
EXCLUDED = 909
SAMPLE SIZE = 1673

L. Billie Joe's third command produces a percentage analysis controlling for RELIG. She sketches down the differences in order to fill in the RELIGIØN column of the blackboard table. A difference of +24 for Protestants and -7 for Catholics shows the subcultural differences to be even more dramatic than in 1960!

YØUR CØMMAND: XTAB:REGIØN,ØCC1,VØTED;%A

REGIØN BY ØCC1 BY VØTED

CØNTRØL: RESPØNDENT'S RESIDENCE
DØWN: RESPØNDENT'S MAIN ØCCUPATIØN - PØL BEH CØDE
ACRØSS: VØTED ØR WØULD HAVE VØTED FØR WHØM IN 68

PERCENTAGES ACRØSS

REGIØN = NISØUT

| | HHH | RMN | TØTAL |
|---|---|---|---|
| WHITE | 44.8% | 55.1% | 303 |
| BLUE | 55.0% | 44.9% | 269 |
| TØTAL | 49.6% | 50.3% | 572 |

REGIØN = SØUTH

| | | | |
|---|---|---|---|
| WHITE | 48.3% | 51.6% | 120 |
| BLUE | 71.7% | 28.2% | 138 |
| TØTAL | 60.8% | 39.1% | 258 |

EXCLUSIØN ANALYSIS:
TABLE TØTAL = 830
EXCLUDED = 843
SAMPLE SIZE = 1673

M

71.7
-48.3
23.4

M. Billie Joe's final COMMAND produces her percentage difference analysis with REGION as a control. She calculates the percent difference for the SOUTH to be 23.4 (the last blackboard table column), makes a quick comparison with the percent difference for the total nation (14.5) and excitedly runs off to prepare for the next class.

## Act 4

Again in the Classroom.

Mass confusion reigns--a student has tripped over several yards of teletype paper being dragged along; a small group is huddled around Billie Joe; and several students are running around asking questions. Professor Knocks enters and order is reestablished with his first statement.

PK: Is Alford still correct on his general proposition?

Ten hands are raised! Ernest Striver blurts out:

ERNEST STRIVER: I got the same answers as Billie Joe! The general proposition of Alford is OK and so are his observations on the subcultural differences as a result of the religious effect. I don't know how to interpret the regional voting patterns--something is wrong!

Ernest Striver is allowed to enter the calculated percentage differences in the blackboard table. His results do indeed agree with Billie Joe's and those of a few other students. A discussion follows and fair agreement with Alford's general proposition is reached. The "religious effect" is more dramatic and is also in agreement with Alford's observations.

PK: Can anyone offer an interpretation of the results on regional voting patterns?

MR. X: Why wasn't REGION dichotomized into New England and not New England?

PK: Astute observation! The standard dichotomies for PRES68 were built by Yankees. However, you could map your own. Back to the question--on the Southern voting patterns?

MISS Z: My friends at A and T say there was an amazing increase of Black votes in 1968!

PK: Could this explain the reversal?

BJW: It certainly could. Black people tend to be concentrated in Blue Collar jobs and Black people tend to vote Democratic. If their voting rates are increasing, this could produce a status effect here in the South where a high proportion of the population is Black.

PK: Is there any way we could test this?

BJW: I'm certain that RACE is a variable in PRES68. We could go back and repeat the analysis with RACE as a control.

PK: Right on! - - - - - -

Professor Knocks is cut off by the rush of students in several directions. Billie Joe leads a group out the door while several students bombard PK with questions as to the location of the teletype, science building, etc.!

- - - - CURTAIN - - - -

Chapter Two

# DESCRIPTION OF THE POISSON SYSTEM

## A. General Introduction

POISSON,* Package Of Instructional Social Surveys Of North Carolina, is a computer system for selective retrieval and analysis of large data files in the social sciences. This system allows searching social science data bases for hypothesis testing within a particular study using specific variables. Selected combinations of 2 to 8 variables can be analyzed using built-in standard dichotomies, standard groupings, or the raw categories of each variable. If desired, the variable categories can be "remapped" to user defined categories. In the 1971 version of POISSON, variable analysis is limited to the discrete mode.

POISSON is available from the IBM 370 model 165 computer at the Triangle Universities Computation Center, TUCC. Four distinct program packages are involved in the system: a tutorial-interactive package (called POISON) which instructs the user in preparing the input data for a run, a batch computational package (called POISSON) which analyzes the data and produces the analyses requested, a data base to the system, and an interactive program package (2X2X2 and WORDS) for three-variable analysis of a user's own data. The use of any of these packages requires no programming knowledge. POISON, 2X2X2, and WORDS are available from the conversational programming system (CPS) at TUCC and are written in PL/I. POISSON and ADDER are batch programs also written in PL/I.

The version of POISSON described in this PRIMER was built around two data bases--PRES68, the Michigan-SRC 1968 presidential survey, and ETHNO, Murdoch's Ethnographic Atlas-Ethnography version. Other data bases are being considered for inclusion in the system. PRES68 and ETHNO were selected with the hope of making social science instruction more inductive while also offering research potential. Additional data bases would enhance the latter purpose.

---

*POISSON was built from the IMPRESS system of Dartmouth College. Permission for the building of the system, the use of IMPRESS codebooks, and the modeling of this primer after the IMPRESS primer has been obtained from Edmund Meyers, Director, Project IMPRESS.

Support of the POISSON system is under the direction of the North Carolina Educational Computing Service, NCECS. Codebooks for PRES68 and ETHNO can be purchased from NCECS. Documentation and user service are obtainable from NCECS, as is assistance in the process of adding a data base: NCECS, P. O. Box 12175, Research Triangle Park, N. C. 27709; phone (919) 549-8291.

## B. The Data Library

The 1971 version of POISSON has two studies on-line: PRES68 (the 1968 election survey--Survey Research Center, University of Michigan) and ETHNO (ethnology version of Murdoch's Ethnographic Atlas). PRES68 has a total unweighted sample N of 1,673 and the POISSON system allows analysis of 109 variables in the data set (this is a subset of the total number of variables available--over 500). ETHNO has a total unweighted sample N of 1,168 and a subset of 91 variables is available from POISSON. Other data sets are being considered for addition to the system and the size of the data sets to be added should be no larger than that of PRES68.

Adding a data set to POISSON necessitates an "inversion" of the original data set from respondent-oriented raw data to a variable-oriented file. The system operates on analysis of variables and the codebooks are built by variables, each of which contains frequencies for each category of the variable. Another requirement for adding a data set is the need to build standard dichotomies and standard groupings for each variable.

Data sets available from Dartmouth College's project IMPRESS are easier to add to POISSON than a new data set. IMPRESS data sets include the variable subsetting, the formation of standard dichotomies and standard groupings, and codebooks giving the frequencies for each category of each variable. The IMPRESS data set itself is not available, however, so that the raw data must be collected and inverted to the format given in the IMPRESS codebooks. Data sets not available at project IMPRESS require, beyond the inversion process, the decision-making for subsetting of available variables and for the formation of standard dichotomies and standard groupings.

To aid in adding a new data set to POISSON, NCECS has made available a package called "ADDER". This package of programs inverts the original data set to the POISSON variable-oriented format once all of the necessary decision

making described in the previous paragraph has been accomplished. Documentation on "ADDER" is available from NCECS.

Codebooks are necessary for adequate use of POISSON, especially when a user needs to map his variable categories to his own liking. PRES68 and ETHNO codebooks are fortunately available from Dartmouth, with minor changes noted in an NCECS addendum to the codebooks. The production of a codebook for a new data set should be considered in the process of adding to POISSON. Almost 30 additional codebooks were available from Dartmouth in 1971.

## C. Access to the System

The IBM 370/165 at TUCC is accessible via several terminal types. Keyboard entry to TUCC is necessary for use of POISON, the interactive version of POISSON. Terminals that have this power include teletypes, IBM 1050's, IBM 2741's, and DATEL's. This same keyboard restriction is necessary to use the programs 2X2X2 and WORDS, the three-variable analysis tutorials. It should be noted that keyboard entry to TUCC is currently limited to low-speed terminals.

POISSON usage is not limited if no keyboard facilities are available. The batch system contains the entire system and terminals without keyboard facilities (such as the IBM 2770 and 2780) can be used for survey analysis. Chapter 4 summarizes the format requirements for card entry to the batch system. Batch-only users are forced to learn the formatting problems (formidable when remapping variables) through which the interactive system, POISON, guides them. An instructor with large classes may find the use of the batch system financially rewarding. At any rate, the interactive system, with its tutorial approach, soon trains a user (after 2 or 3 runs) to go directly to the batch system without using the interactive system again. All users will find this approach financially rewarding.

Accessing the three variable analysis tutorial programs is by keyboard terminal only, since both 2X2X2 and WORDS are in the CPS system (interactive PL/I). These programs allow three-variable cross-tabulation with analysis by gamma and have the advantage of allowing a user to submit his own data for analysis (the POISSON system is restricted to mounted data sets). These programs are described in Chapter Six.

Chapter Three

# A WALK THROUGH THE INTERACTIVE VERSION OF POISSON

## Introduction

The scenario of Chapter One assumed a user experienced with POISSON. While the interactive version of POISSON supplies a tutorial approach to produce experienced users, it is advantageous to document this interactive system in detail for local trouble-shooting support. Further, the options for assistance in formulating a question to a data set are rich in the interactive system and an overview is almost essential. If an interested user has no interactive facilities (low speed, keyboard terminal), he must read this chapter since it outlines the logic necessary for analysis. However, it is repeated here that the use of POISSON is in no way precluded for users without interactive facilities. The 1971 interactive version merely formulates the card format for running the system and does no analysis itself.

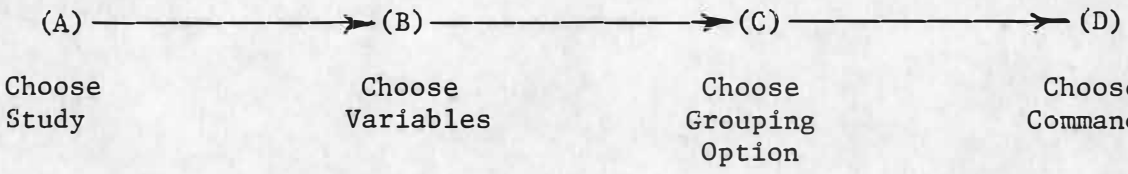
Once the interactive program is called, as shown in the scenario of Chapter One, the program first types "DO YOU WANT USER'S INSTRUCTIONS?" These instructions call attention to two levels of assistance available at every stage in which the computer asks for input. The user may type in:

EXPLAIN (for a brief explanation)

DETAIL (for a longer explanation)

It is suggested that the beginner type, "YES" for user's instructions and "DETAIL" for every initial new request from the computer.

The skeleton of the interactive version of POISSON looks like this:




The process will be stepped through from left to right just as the user sees it. Each stage is equipped with instructions if the user types in "EXPLAIN" or "DETAIL".

(A) Choosing a Study

When the system types, "ENTER NAME ØF SURVEY", it is asking you to type in the name of the specific survey with which you will be working (only one survey at a time).

ENTER NAME ØF SURVEY
?PRES68

If you just want a list of available survey names, type LIST. This request will produce the codenames of all available surveys on line. After these names are given, the previous question is repeated by the computer:

ENTER NAME ØF SURVEY

This repetition of the question will occur after all responses such as EXPLAIN, DETAIL, LIST, and DESCRIBE (see below) which ask for assistance before answering the question itself. An improper spelling of the survey codename or the typing of a name not available will cause an error message and again a repeat of the question.

(B) Choosing Variables

Having located the survey requested, the computer then requests:

ENTER NAMES ØF 2 TØ 8 VARIABLES TØ BE RUN (CØMMAS)--NØ BLANKS PLEASE
?

Every variable in every POISSON survey has a specific name (e.g., MARITL, REPUBS, RØCKY). They are chosen to be meaningful (you can guess that MARITL has something to do with marriage, REPUBS with Republicans, and RØCKY with Rockefeller). Sometimes, though, they are cryptic; (would one guess that NEGATT refers to the respondent's estimate of the opinion of Negroes in his area regarding desegragation?).

A proper response to this question would be:

?MARITL,REPUBS,RØCKY,SØUTH,WHØJØB,WAR1

Note the formatting with respect to commas and blanks. A request for assistance by a response of EXPLAIN or DETAIL would expose five specific approaches to getting more information about the variable in the survey:

(1) Type LIST and the system will type out the names of all variables in the study. Warning: Up to 109 variable names is a lot of time-consuming output. A careful inspection of the codebook could

eliminate this.

(2) Type DESCRIBE:MARITL,REPUBS,RØCKY and the system will type out a capsule (50 characters or less) description of these three variables.

(3) Type DESCRIBE:MARITL,REPUBS,RØCKY;SD (note the absence of blanks and the specific punctuation) and POISON will produce the short description and category labels with accompanying frequencies for the standard dichotomy (SD) of each variable.

(4) Type DESCRIBE:MARITL,REPUBS,RØCKY;SG and the items of (3) above will be produced for the standard groupings (SG) of each variable.

(5) Type DESCRIBE:MARITL,REPUBS,RØCKY;ALL and POISON will give all of (2) and (3) above as well as the total number of categories for the raw data of each variable.

All of these aids in choosing variables will be followed by a repeat of the request for the names of the variables to be chosen.

(C) Choosing the Grouping Option

The variables having been selected, the next computer-generated request will be:

```
ENTER GRØUPING ØPTIØN
?
```

Typing EXPLAIN or DETAIL will output an explanation of six options available to answer this request, all having to do with how the categories of each variable are to be grouped for analysis:

(1) IND — the grouping of each variable will be entered individually. This option is one way to group each variable in a different manner (often desirable), these manners being options (3), (4), (5), and (6) following.

(2) DESCRIBE — prints a short description of each variable label (handy for cryptic labels which make no sense), prints the SD and SG categories and frequencies, then asks for one of the following grouping options.

(3) MAP — allows user to regroup the raw categories to his own choosing and therefore rejects the raw, SD, and SG groups of the codebook.

(4) RAW gives the user access to the categories as they were in the original data.

(5) SG stands for "Standard Groups"; see codebooks for these built in groupings of three or more categories for each variable.

(6) SD stands for "Standard Dichotomy"; choosing this option accepts the two codebook SD categories together with a third group of leftovers ("DØNT KNØW, NØ ANSWER, AND ØTHERS.")

The inexperienced user will probably feel most comfortable using SD or SG and an answer to the request:

```
ENTER GRØUPING ØPTIØN
?
```

of SD or SG will establish the chosen grouping for all variables previously selected. Eventually, the user may try to MAP (regroup) the raw categories in order to test his chosen hypothesis.

The use of the MAP option is somewhat complex and is specifically illustrated for the variable KIDS of survey PRES68. In the codebook, KIDS appears exactly as in the following table:

KIDS P. 2A DECK #1 COL. 53 - 53

ARE THERE ANY CHILDREN UNDER 18 YEARS ØLD IN THIS FAMILY? (HØW MANY?)

| FREQ. | % | SD LABEL | SG LABEL | IMP #'S | ØRIG #'S | |
|---|---|---|---|---|---|---|
| 858 | 51.2 | NØNE | NØNE | 0 | 0 | NØNE UNDER 18; INAP., CØDED 2 IN P.2 |
| 243 | 14.5 | KIDS | ØNE | 1 | 1 | ØNE |
| 242 | 14.4 | KIDS | TWØ | 2 | 2 | TWØ |
| 159 | 9.5 | KIDS | THREE | 3 | 3 | THREE |
| 72 | 4.3 | KIDS | 4ØRMØR | 4 | 4 | FØUR |
| 44 | 2.6 | KIDS | 4ØRMØR | 5 | 5 | FIVE |
| 26 | 1.5 | KIDS | 4ØRMØR | 6 | 6 | SIX |
| 15 | 0.8 | KIDS | 4ØRMØR | 7 | 7 | SEVEN |
| 9 | 0.5 | KIDS | 4ØRMØR | 8 | 8 | EIGHT ØR MØRE |
| 5 | 0.2 | NA | NA | 9 | 9 | NA |

The column headed "SD LABEL" indicates that KIDS has been dichotomized into a NØNE group (no kids under 18) and a KIDS group for people who have 1 or more kids under 18. The third group, NA, is a category for those respondents who did not answer this question. The column headed "SG LABEL" gives the same type of information for Standard Groups.

Assume the user wants to group KIDS into three categories:

(1) respondents claiming one or two children,

(2) respondents claiming three or more children,

(3) respondents not answering plus those claiming no children in the family.

After typing MAP, the POISON system will respond:

```
'KIDS' LØCATED - ANY CHILDREN UNDER 18 IN FAMILY?
HØW MANY NEW CATEGØRIES?
?
```

Type in the number 3 for the three new groups chosen. The computer will then ask:

```
GRØUP 1?
?
```

The answer must be given as a name of the new category (6 characters or less) followed by the IMPRESS numbers of the categories to be included in this group (from column headed IMP #'S). Separate the name and category numbers by commas (no blank) as follows:

```
GRØUP 1?
?1ØR2,1,2
```

The computer will then type:

```
GRØUP 2?
?
```

The second group chosen, 3ØRMØRE, would include six categories and can be entered as follows:

```
GRØUP 2?
?3ØRMØR,3-8
```

The hyphen saves much typing for larger blocks of categories. As usual, no blanks can be used. After the next computer request:

```
GRØUP 3?
?
```

the user must be sure to include all the remaining raw categories for this group. Since two categories remain in this example, 0 and 9, the response would be:

```
GRØUP 3?
?NØNENA,0,9
```

The computer then types out:

```
ENTER EXCLUSIØN CØDE
?
```

The user is hereby requested to clarify how many of his new categories should be included in his tables and calculations. A numerical code is available to answer this request:

0 - for no exclusions.
1 - for exclusion of the highest numbered group.
2 - for exclusion of the two highest numbered groups.
(etc.)

In the sample regrouping, the analyst is assumed to be interested in the two new groupings, 1ØR2 and 3ØRMØR, and would like to exclude the new grouping, NØNENA. The response should be:

```
ENTER EXCLUSIØN CØDE
?1
```

By typing a 1, the user excludes the highest numbered group, group 3. Note that the formation of the new groupings should have this code in mind since he can only exclude the highest numbered group(s).

As was mentioned above, a user can give a blanket grouping option to cover all of the variables he has chosen. However, the more usual case will require different grouping options for the variables. If the option IND is typed, each variable will be treated in turn by the computer. There is an alternate method of making exceptions to a blanket option. For example:

```
ENTER GRØUPING ØPTIØN
?SG,EXCEPT,EDUC,SD,KIDS,MAP
```

Again note the punctuation and the necessary absence of blanks. In this example, all variables will be treated by the "Standard Grouping" option except EDUC which will be treated as a "Standard Dichotomy" and KIDS, which will be regrouped or "MAP"ed.

A final note on the grouping of categories. The order in which new groups are formed should be considered in advance if the new groups are to represent an ordered grouping. The new group with the highest ordering should be entered first. Ordering therefore goes from high to low as the numbers of the groups increase. Remember, HIGHEST GROUP FIRST.

(D) Choosing Commands

Having specified the grouping of each variable, the user now enters the analytic stage. Two analysis commands are available in the 1971 version of POISSON:

MARG: gives the marginal frequencies for each category of the variables requested.

XTAB: means "cross-tabulation" and gives tables in which each variable is run against all of the other chosen variables.

In the interactive version of POISSON, the computer generates the following question once the grouping options have been determined:

```
ENTER COMMAND
?
```

A command for analysis ultimately requires the command, the variables to be treated, and the statistical option(s) requested for the analysis. The fastest way through POISON is to enter all of this information at once as in the following example:

```
ENTER COMMAND
?XTAB:SEX,EDUC;%A,CHI
```

This example requests cross-tabulation of SEX and EDUC with the statistical options of %A and CHI (see below for details on the statistical options). Note the necessary punctuation and the absence of blanks. A colon follows the command and a semicolon separates the variables from the statistical options.

It is not necessary to supply all this information at once and the novice is better off just entering a command as follows:

```
ENTER COMMAND
?XTAB
```

The computer will then enter requests for the variables and statistical options

in turn:

```
ENTER VARIABLES (SEPARATE LIST BY CØMMAS)
?
```

For three-variable analysis, the order of entry should be:

CONTROL, INDEPENDENT, DEPENDENT

For example:

```
ENTER VARIABLES (SEPARATE LIST BY CØMMAS)
?RELIG,ØCC1,VØTED
```

In this example RELIG is the control variable, ØCC1 the independent variable, and VØTED the dependent variable. Note that the variables entered must be chosen from those specified at the beginning of the POISON run.

Once the variables are entered, a request is generated for the statistical options desired:

```
ENTER ØPTIØNS (SEPARATE LIST BY CØMMAS)
?
```

The available options (with their code abbreviations) in the 1971 version of POISSON are as follows:

| | | |
|---|---|---|
| CHI | - | Chi Square Statistics |
| FREQ | - | Frequency Table |
| %A | - | Percentages Across Table |
| %D | - | Percentages Down Table |
| %TAB | - | Percentages of Total Table |
| NØTAB | - | No Table |
| SELECT | - | Use One Group of a Variable Only |
| REJECT | - | Reject One Group of a Variable Only |
| GAMMA | - | Gamma Statistic (Yules Q if the table is only 2x2x2) |
| EXP | - | Table of Expected Frequencies |
| DELTA | - | Table of Frequencies minus expected frequencies. |

An example of a response to the request for options would be:

```
ENTER ØPTIØNS (SEPARATE LIST BY CØMMAS)
?%A,CHI
```

Note the absence of blanks and the necessary punctuation.

Once the command, variables, and statistical options have been entered, all at once or in steps as has been illustrated immediately above, the analysis

stage has been fully formulated. At this point, in the 1971 version of POISSON, the interactive version can go no further in producing analysis and merely prints out the following message:

```
YØUR CØMMAND, VARIABLES, AND ØPTIØNS HAVE BEEN ACCEPTED.
WØULD YØU LIKE TØ ENTER ANØTHER CØMMAND (YES ØR NØ)?
?
```

This request allows the user to enter up to a total of five commands so that he may do several analyses of his variables at one crack. If the answer "YES" is typed, he would repeat the procedure for entering a command, the variables, and the statistical options after the computer generates the requests for this information. A "NØ" answer would produce the following message, which is a sample final output of the 1971 interactive version of POISSON:

```
        THE FØLLØWING IS A PRINT-ØUT ØF YØUR CØNTRØL CARDS FØR
    A 'PØISSØN' RUN.  THE FIRST TWØ LINES ARE CARD CØLUMNS
    AND NØT TØ BE PUNCHED.

         1    1    2    2    3    3    4    4    5    5    6    6    7
....5....0....5....0....5....0....5....0....5....0....5....0....5....0
ID          PRES68     REGIØN,RELIG,ØCC1,VØTED
GRØUPING    IND        REGIØN=SD,RELIG=MAP,ØCC1=SD,VØTED=MAP
MAP         RELIG,1    PRØT,0-4;CATH,5;ØTHNA,6-13
MAP         VØTED,1    HHH,0,1,14,15;RMN,2,3,16,17;ØTHNA,4-13,18-34
CØMMAND     MARG:RELIG,VØTED
CØMMAND     XTAB:ØCC1,VØTED;%A
CØMMAND     XTAB:RELIG,ØCC1,VØTED;%A
CØMMAND     XTAB:REGIØN,ØCC1,VØTED;%A
```

This output example is the same as the one in the scenario of Chapter One. The first two rows of print-out represent punched card column numbers as a reference for key-punching. Eight lines of output show the exact format for punched cards to do four commands on four variables, two of which variables have been regrouped, the other two treated as "Standard Dichotomies".

"LØGØUT" procedures then terminate the running of this interactive version and the actual calculations are left to the batch system (see Chapter Four). The format of the punched cards soon becomes familiar (note the importance of columns 1, 11, and 21) so that the user may go directly to the batch system after learning what it is all about through the interactive version. The sacrifice of immediate feedback of answers in the interactive (time-sharing) mode is offset by the cost effectiveness of using a batch

system. Inclusion of all the calculations, data bases, etc., in the interactive version is not possible at TUCC at the writing of this primer due to storage limitations in the conversational programming system (CPS).

Chapter Four

# THE BATCH VERSION OF POISSON

## A. Introduction

The batch version, POISSON, contains the complete analysis system necessary to complement the interactive version described in Chapter Three. Formatting punched cards (or keyboard entries) for entering a run in the batch mode, described in this chapter, is done for the user in POISON. To understand the references to the contents of each formatted "card", a knowledge of the process for setting up an analysis, as described in Chapter Three, is essential.

A batch run of POISSON can be described in terms of "control cards". Each of the five possible "control cards" represents a line of data on a keyboard entry terminal or a single punched card. These "control cards" are sandwiched in between another set of cards, called "JCL" - Job Control Language - cards, the latter being necessary to contact the POISSON system at TUCC. "Control cards" are described in section B and "JCL" cards in section C below.

## B. Control Cards for POISSON

(1) The ID Card:

Each batch run must contain one ID card and it must be the first control card. Its required format is:

Columns

| 1 | 11 | 21 |
|---|---|---|
| ID | survey name | names of variables to be included in run, separated by commas |

EXAMPLES: (Column numbers are references and are not to be punched.)

```
         1    1    2    2    3    3    4    4    5    5    6    6    7
1...5....0....5....0....5....0....5....0....5....0....5....0....5....0
```

```
ID        PRES68     REGIØN,RELIG,ØCC1,VØTED

           or

ID        ETHNØ      TABØØS,TERCUS,SLAVE1
```

(2) The GRØUPING card

Each batch run must also contain one GRØUPING card and it must be the second control card. This card specifies the grouping options for each variable used in the run. Two formats are possible for this card--the first being used for a blanket grouping option requested for each variable (with exceptions noted) and the second for the IND option, in which each variable is grouped separately. The format for this control card is:

Columns

```
1          11          21
GRØUPING   Grouping    Exceptions to the Grouping Option
           option
              or                        or
           IND         Grouping options for each variable
```

Examples of the two formats are (column numbers are references and are not to be punched):

```
          1    1    2    2    3    3    4    4    5    5    6    6    7
1...5....0....5....0....5....0....5....0....5....0....5....0....5....0
GRØUPING  SD        RELIG=MAP,VØTED=MAP

                or

GRØUPING  IND       IABØØS=MAP,TERCUS=SD,SLAVE1=SD
```

In the first example only the variables which are exceptions to the blanket SD grouping options are specified in column 21 onwards. The second example calls for the IND option and each variable given on the ID control card must be specified for its own grouping option; i.e., TABØØS, TERCUS, and SLAVE1 represent the complete list of variables on the ID card.

If the user needs more space on the GRØUPING control card, he should end his control card at a comma (the end of the grouping option requested for any one variable) and start a new card in column 21 continuing his specification of grouping options. For example:

```
          1    1    2    2    3    3    4    4    5    5    6    6    7
1...5....0....5....0....5....0....5....0....5....0....5....0....5....0

GRØUPING  IND       KIDS=MAP,VØTED=SD,ØCC1=SG,RØCKY=RAW,
                    RFK=SD,ATTHHH=SD,ATTNIX=SD,ATTLBJ=SD
```

(3) The MAP card:

For every variable that uses MAP as a grouping option, a MAP card is required which contains the regrouping information necessary for a run (if no variables use the MAP option, no MAP card will be used). The MAP cards (and the RAW cards described in (4) below) always follow the GRØUPING card (the order of a series of MAP and RAW cards is immaterial). The required format is:

| Columns 1 | 11 | 21 |
|---|---|---|
| MAP | Name of Variable and Exclusion Code | The new grouping names, each accompanied by appropriate codebook numbers for the raw categories to be combined. |

Example: (punctuation must be precise)

```
        1    1    2    2    3    3    4    4    5    5    6    6    7
1...5....0....5....0....5....0....5....0....5....0....5....0....5....0

MAP       VØTED,1   HHH,0,1,14,15;RMN,2,3,16,17;ØTHNA,4-13,18-34
```

In this example, the variable VØTED has been regrouped into three new groups: HHH (containing the raw categories 0, 1, 14, and 15), RMN (raw categories 2, 3, 16, and 17), and ØTHNA (with raw categories 4 through 13 and 18 through 34). Note that the exclusion code of 1 is established after the variable name.

The use of a dash between the extremes of a large group of raw categories eases the punching process. Nevertheless, there will be cases in which a card (or a line) is not long enough for all the information and continuation can be done on a new card or line starting at column 21. The break at the end of any one card (or line) must occur after a semicolon.

Example:

```
        1    1    2    2    3    3    4    4    5    5    6    6    7
1...5....0....5....0....5....0....5....0....5....0....5....0....5....0

MAP       ØCC1,1    PRØF,0-199;MGRS,200-299;CLERKS,300-349;
                    SALES,350-399;CRAFTS,400-499;ØPS,500-599;
                    ØTHNA,600-998
```

(4) The Raw card:

For every variable that uses RAW as a grouping option, one RAW card is

necessary. The RAW card(s) contains user-supplied names of the raw categories of the variable requested and the exclusion code for these categories. This card can take two basic formats as follows:

Columns

1 11 21

RAW Variable Name and Exclusion Code | Names of Raw Categories | or | Number

Examples:

```
         1    1    2    2    3    3    4    4    5    5    6    6    7
1...5....0....5....0....5....0....5....0....5....0....5....0....5....0
```

RAW KIDS,2 ZER,ØNE,TWØ,THREE,FØUR,FIV,SIX,SEV,EIGHT,ØTHNA

or

RAW RELIG,1 NUMBER

The establishment of names for the raw categories is allowed only when there are 10 or less raw categories. If there are more than 10 raw categories, they will be numbered and the word NUMBER is punched in columns 21 through 26.

The exclusion code format is identical to that for the MAP card. However, a default exclusion code is available for each variable and, if no comma and number are entered after the variable name, this default exclusion code is invoked. For example:

```
         1    1    2    2    3    3    4    4    5    5    6    6    7
1...5....0....5....0....5....0....5....0....5....0....5....0....5....0
```

RAW REGIØN NØRTH,SØUTH,EAST,WEST

(5) The CØMMAND card:

Each batch run must contain at least one CØMMAND card and can have as many of these cards as desired. This card contains the CØMMAND desired (MARG or XTAB), the variables to be used in this command (in the following order if only three variables--CONTROL, INDEPENDENT, and DEPENDENT variables), and the statistical options requested for the command. The format of this card

is as follows:

Columns:

1 11

CØMMAND Type of command, variables, and statistical options

Examples:

```
         1    1    2    2    3    3    4    4    5    5    6    6    7
1...5....0....5....0....5....0....5....0....5....0....5....0....5....0....
```

```
CØMMAND   MARG:RELIG,VØTED
CØMMAND   XTAB:ØCC1,VØTED;%A
CØMMAND   XTAB:RELIG,ØCC1,VØTED;%A
CØMMAND   XTAB:REGIØN,ØCC1,VØTED;%A,GAMMA
```

Note the colon between the command and the variables, the semicolon between the variables and the statistical options, and the commas between variables and between statistical options (when there are more than one).

## C. JCL Cards

A batch run must tell the computer where to bill you and how to access POISSON itself. Five cards are necessary and they are always of the following form (note the sandwiching in of the control cards described in section (B) above):

```
//jobname  JØB  xxx.yyy.zzz,name
//  EXEC   PØISSØN
//SYSIN  DD  *
       Control Cards
/*
//
```

The small letters in the first card, the JØB card, are user-supplied. The "jobname" is any name of eight characters or less to identify your run. The "xxx.yyy.zzz" is an account code available from whoever manages your terminal. Your name, "name", is another user-supplied item. All of the rest remains exactly as in the above example. Sandwich in the control cards, get this deck (or keyboard lines) into your terminal and you're all set.

Chapter Five

## SAMPLE RUNS - BATCH VERSION

This chapter is devoted to illustrating output from the batch version, POISSON, for the statistical options available in the system. Under the CØMMAND of XTAB (cross-tabulation), the following statistical options are available in the 1971 version of the system:

| | |
|---|---|
| CHI | Chi Square Statistic |
| FREQ | Frequency Table |
| %A | Percentages Across Table |
| %D | Percentages Down Table |
| %TAB | Percentages-of-Total Table |
| NØTAB | No Table |
| SELECT | Use One Group of a Variable Only |
| REJECT | Reject One Group of a Variable Only |
| GAMMA | Gamma Statistic (Yule's Q for a 2x2x2 table) |
| EXP | Table of Expected Frequencies |
| DELTA | Table of Frequencies minus Expected Frequencies |

These options will be illustrated with output from a batch run with a series of CØMMANDS run on one set of variables.

CONTROL CARDS:

```
         1    1    2    2    3    3    4    4    5    5    6    6    7
1...5....0....5....0....5....0....5....0....5....0....5....0....5....0

ID         PRES68     REGIØN,RELIG,KIDS
GRØUPING   IND        REGIØN=SD,RELIG=MAP,KIDS=SG
MAP        RELIG,1    PRØTS,0-4;CATH,5;ØTHER,6-13
CØMMAND    MARG:RELIG
CØMMAND    XTAB:RELIG,KIDS;CHI,FREQ,%A,%TAB,GAMMA
CØMMAND    XTAB:REGIØN,RELIG,KIDS;EXP,DELTA,GAMMA
CØMMAND    XTAB:RELIG,KIDS;FREQ,SELECT(KIDS=THREE)
```

Three variables from PRES68 are used to illustrate the statistical options: REGIØN (standard dichotomy), RELIG (MAP), and KIDS (standard grouping). The first command illustrates the use of MARG to check the resulting marginals for the variable RELIG after this variable was regrouped ("MAP'ed"). XTAB is used in the remaining commands to give sample output for the statistical options.

Commands two and four deal with two-variable analysis while command three utilizes a control variable. The REJECT option is not illustrated since its format is identical to that of the SELECT option seen in the fourth command.

Output for this single run with four commands is as follows:

```
ID          PRES68      ØCC1,REGIØN,RELIG,KIDS
GRØUPING    IND         ØCC1=SD,REGIØN=SD,RELIG=MAP,KIDS=SG
MAP         RELIG,1     PRØTS,0-4;CATH,5;ØTHER,6-13

---------------
YØUR CØMMAND:  MARG:RELIG

                MARGINAL FREQUENCIES

RELIG   =  RELIGIØUS PREFERENCE

PRØTS           1167   69.8%
CATH             348   20.8%
ØTHER            158    9.4% EXCLUDED
```

YØUR CØMMAND: XTAB:RELIG,KIDS;CHI,FREQ,%A,%D,%TAB,GAMMA

RELIG BY KIDS

DØWN: RELIGIØUS PREFERENCE
ACRØSS: NUMBER OF CHILDREN UNDER 18

FREQUENCY TABLE

| | NØNE | ØNE | TWØ | THREE | 4ØRMØR | TØTAL |
|---|---|---|---|---|---|---|
| PRØTS | 619 | 172 | 173 | 101 | 100 | 1165 |
| CATH | 159 | 53 | 43 | 37 | 55 | 347 |
| TØTAL | 778 | 225 | 216 | 138 | 155 | 1512 |

PERCENTAGES ACRØSS

| | NØNE | ØNE | TWØ | THREE | 4ØRMØR | TØTAL |
|---|---|---|---|---|---|---|
| PRØTS | 53.1% | 14.8% | 14.8% | 8.7% | 8.6% | 1165 |
| CATH | 45.8% | 15.3% | 12.4% | 10.7% | 15.9% | 347 |
| TØTAL | 51.5% | 14.9% | 14.3% | 9.1% | 10.3% | 1512 |

PERCENTAGES DØWN

| | | | | | | |
|---|---|---|---|---|---|---|
| PRØTS | 79.6% | 76.4% | 80.1% | 73.2% | 64.5% | 77.1% |
| CATH | 20.4% | 23.6% | 19.9% | 26.8% | 35.5% | 22.9% |
| TØTAL | 778 | 225 | 216 | 138 | 155 | 1512 |

PERCENTAGES ØF TØTAL

| | NØNE | ØNE | TWØ | THREE | 4ØRMØR | TØTAL |
|---|---|---|---|---|---|---|
| PRØTS | 40.9% | 11.4% | 11.4% | 6.7% | 6.6% | 77.1% |
| CATH | 10.5% | 3.5% | 2.8% | 2.4% | 3.6% | 22.9% |
| TØTAL | 51.5% | 14.9% | 14.3% | 9.1% | 10.3% | 1512 |

CHI SQUARE = 18.890 WITH 4 D.F.
THE PRØBABILITY ØF CHI SQUARE IS < .001

GAMMA

G = -0.150 (RELIG = PRØTS ) TENDS WITH (KIDS = 4ØRMØR)

EXCLUSIØN ANALYSIS:
TABLE TØTAL = 1512
EXCLUDED = 161
SAMPLE SIZE = 1673

---

YØUR CØMMAND: XTAB:REGIØN,RELIG,KIDS;EXP,DELTA,GAMMA

REGIØN BY RELIG BY KIDS

CØNTRØL1: RESPØNDENT'S RESIDENCE
DØWN: RELIGIØUS PREFERENCE
ACRØSS: NUMBER ØF CHILDREN UNDER 18

REGIØN = NTSØUT

EXPECTED CELL FREQUENCIES

| | NØNE | ØNE | TWØ | THREE | 4ØRMØR | TØTAL |
|---|---|---|---|---|---|---|
| PRØTS | 339 | 100 | 95 | 66 | 72 | 673 |
| CATH | 157 | 46 | 44 | 30 | 34 | 311 |
| TØTAL | 496 | 146 | 139 | 96 | 106 | 983 |

ØBSERVED - EXPECTED FREQUENCY TABLE

| | NØNE | ØNE | TWØ | THREE | 4ØRMØR |
|---|---|---|---|---|---|
| PRØTS | 14 | -3 | 4 | -2 | -13 |
| CATH | -14 | 3 | -4 | 2 | 13 |

GAMMA

G = -0.127 (RELIG = PRØTS) TENDS WITH (KIDS = 4ØRMØR)

REGIØN = SØUTH

EXPECTED CELL FREQUENCIES

| | NØNE | ØNE | TWØ | THREE | 4ØRMØR | TØTAL |
|---|---|---|---|---|---|---|
| PRØTS | 263 | 74 | 72 | 39 | 46 | 493 |
| CATH | 19 | 5 | 5 | 3 | 3 | 36 |
| TØTAL | 282 | 79 | 77 | 42 | 49 | 529 |

ØBSERVED - EXPECTED FREQUENCY TABLE

| | NØNE | ØNE | TWØ | THREE | 4ØRMØR |
|---|---|---|---|---|---|
| PRØTS | 3 | 1 | 2 | -2 | -5 |
| CATH | -3 | -1 | -2 | 2 | 5 |

GAMMA

G = -0.249 (RELIG = PRØTS ) TENDS WITH (KIDS = 4ØRMØR)

PARTIAL = -0.136 (RELIG = PRØTS ) TENDS WITH (KIDS = 4ØRMØR)

ZERO ORDER = -0.150

(RELIG = PRØTS ) TENDS WITH (KIDS = 4ØRMØR)

EXCLUSIØN ANALYSIS:
TABLE TØTAL = 1512
EXCLUDED = 161
SAMPLE SIZE = 1673

```
---------------
YØUR CØMMAND:    XTAB:RELIG,KIDS;FREQ,SELECT(KIDS=THREE)

RELIG   BY  KIDS

DØWN:       RELIGIØUS PREFERENCE
ACRØSS:     NUMBER ØF CHILDREN UNDER 18
```

FREQUENCY TABLE

| | NØNE | ØNE | TWØ | THREE | 4ØRMØR | TØTAL |
|---|---|---|---|---|---|---|
| PRØTS | 0 | 0 | 0 | 101 | 0 | 101 |
| CATH | 0 | 0 | 0 | 37 | 0 | 37 |
| TØTAL | 0 | 0 | 0 | 138 | 0 | 138 |

```
EXCLUSIØN ANALYSIS:
TABLE TØTAL  =   138
EXCLUDED     =  1535
SAMPLE SIZE  =  1673
```

Chapter Six

# THREE-VARIABLE ANALYSIS PROGRAMS IN THE TUTORIAL-INTERACTIVE MODE

Two programs dealing with three-variable analysis are available in the CPS system and thus in the interactive mode. These programs, called 2X2X2 and WORDS, allow a user to enter his own data for a 2x2x2 contingency table and analyze this data using the gamma statistic (called Yule's Q when the contingency table is limited in size to 2x2x2). Both programs allow the "decomposition" of the gamma relationship between the dependent and independent variables in terms of the control variable so that the control variable can be determined as "explaining," having "no effect" on, "suppressing," or merely partially affecting, the relationship between the dependent and independent variables. The program 2X2X2 performs the calculations and tabulates the results while WORDS aids in preparing a verbal report of the analysis of the relationship.

The two programs have two purposes distinct from those of the POISSON system. They perform a tutorial task in teaching the understanding of gamma as used with a 2x2x2 contingency table. Secondly, they provide the user with a method of analysing published results, in both the popular media and in scientific journals, if the variable reported can be easily broken into standard dichotomies and if the frequencies of the observations are available. Specific examples of both purposes are furnished in the following sample runs of 2X2X2 and WORDS.

## A. An "Explanation" Variable

The sociologist Jan Hajda surveyed the reading habits in a probability sample of 1,850 women in Baltimore, Maryland. He observed a low negative correlation between age (under 45 vs. 45 or older) and bookreading (the number of books the respondent claimed to have read in the previous year). Suspecting that this correlation may be due to another variable, Hajda drew on education as a possibility. The historical analysis will be followed below using 2X2X2, WORDS, and Hajda's raw data. This illustration will demonstrate the tutorial purpose of the programs as well as the research-checking capacity as an educational tool.

A
```
SERVICE?
CPS
?xxx.yyy.zzz,name
GØØD AFTERNØØN; USER 04; TIME 13:28:29  10/27/71
```

B
```
?LØAD(2X2X2)PUBLIC
?XEQ
DØ YØU WANT USER'S INSTRUCTIØNS?  'Y' ØR 'N'
?N
```

C
```
SUPPLY LABELS ØR ACCEPT X,Y,&T ('LABELS' ØR 'X?I')
?LABELS

NØW YØUR LABELS
FØR DEPENDENT VARIABLE, Y:
?READ,HIGH,LØW

FØR INDEPENDENT VARIABLE, X:
?AGE,45UP,44DW

FØR TEST VARIABLE, T:
?EDUC,HIGH,LESS
DID YØU ENTER YØUR VARIABLES CØRRECTLY?
TYPE 'Y' IF YØU DID, 'N' IF YØU WANT TØ RE-DØ THEM
?Y
```

D
```
ENTER A,B,C,D,E,F,G,H
A
?215,263,373,453,335,54,133,24
```

RAW DATA

| | | /READ/ LØW | HIGH | |
|---|---|---|---|---|
| EDUC | AGE | | | |
| HIGH | 45UP | 215 | 263 | 478 |
| | 44DW | 373 | 453 | 826 |
| | | 588 | 716 | |
| LESS | 45UP | 335 | 54 | 389 |
| | 44DW | 133 | 24 | 157 |
| | | 468 | 78 | |
| | | | | 1850 |

A. User dials into the CPS system and is recognized from his account code.

B. He loads the program 2X2X2 and calls for the execution of the program (two distinct steps indicated by the initial question marks). The computer asks the user if he needs instructions and he answers NØ.

C. Then 2X2X2 asks for variable labels. Hajda's variables are entered as user supplied codes--READ is the code for the dependent variable "book reading"--the dichotomy categories being given as HIGH and LØW; AGE (the independent variable) is dichotomized into 45UP (45 and older) vs 44DW (44 and below); and the control variable education (EDUC) is dichotomized into HIGH (high school education and above) and LESS (less then full high school education). Note that the label for the higher order of each dichotomy is entered first and that the punctuation is critical. The computer allows a check of these label entries and will ask for them again if the user catches an error in his labels.

D. Hajda's raw data is entered into the computer. User's instructions, not called upon in this run, show the relation between the table entries and the A through H items requested. The raw data is entered in one line (they could be entered one at a time) and the computer echoes out the raw data 2X2X2 table (with marginals). If incorrectly entered, 2X2X2 will then allow reentry of the data after the check.

D

```
IS THIS TABLE CØRRECT?  ('Y' ØR 'N')
?Y
```

E

```
DØ YØU WANT THE MARGINAL PERCENTAGES?
?Y
```

MARGINAL PERCENTAGES

| | AGE PERCENT | READ PERCENT | EDUC PERCENT |
|---|---|---|---|
| UPPER LIMIT (.95) | 49.1 | 45.2 | 72.6 |
| SAMPLE VALVE | 46.9 | 42.9 | 70.5 |
| LØWER LIMIT (.95) | 44.6 | 40.7 | 68.4 |

F

```
DØ YØU WANT A PERCENTAGE TABLE?
?Y
```

PERCENT HIGH

AGE

| | 44DW | 45UP |
|---|---|---|
| HIGH | 54.8 (826) | 55.0 (478) |
| EDUC | | |
| LESS | 15.3 (157) | 13.9 (389) |

G

```
DØ YØU WANT ZERØ ØRDER Q's
?Y
```

ZERØ ØRDER Q LIMITS AND SIGNIFICANCE

| | Q(AGE,READ) | Q(EDUC,AGE) | Q(EDUC,READ) |
|---|---|---|---|
| UPPER LIMIT (.95) | -.152 | -.554 | .815 |
| SAMPLE VALUE | -.240 | -.620 | .759 |
| LØWER LIMIT (.95) | -.328 | -.687 | .703 |

E. Now the analysis, step by step. The computer asks if a MARGINAL PERCENTAGE TABLE is desired. An answer of Y produces this table, a relevant consideration before analysis since the table provides some degree of assurance of the validity of analysis based on the size of the dichotomies. The percentages and upper and lower limits indicate a reasonable cell size for each dichotomy and the analysis continues.

F. The program then asks if a PERCENTAGE TABLE is desired. Again, an answer of Y produces the table. The percentage high on READ is given for EDUC vs AGE. A good analyst would quickly note that the percentage high on reading is almost identical for both the higher education group and the lower education group and that the percentage difference in reading for both age groups is about constant. A good indication of the effect of education on reading habits regardless of the age group!

G. Now the gamma analysis! (Yules Q is the gamma statistic for a 2x2x2 table and the programs use Q throughout for this 2x2x2 analysis.) An answer of Y produces the zero order Q's--each combination of variables without a control. Note the low negative association between AGE and READ discovered by Hajda (-.240). Also note the strong negative association between EDUC and AGE and the strong positive association between EDUC and READ. This large spread (with opposite signs) indicates a strong possibility of EDUC "explaining" the association between AGE and READ.

DØ YØU WANT THE DECØMPØSITIØN ØF Q?
?Y

DECØMPØSITIØN ØF Q

H

| | PAIR (AGE,READ) | (AGE,EDUC) | (EDUC,READ) |
|---|---|---|---|
| Q ZERØ ØRDER | -.240 | -.620 | .759 |
| Q 3RD ITEM TIED | 0.000 | -.620 | .757 |
| WEIGHT 3RD ITEM TIED | .498 | .465 | .443 |
| Q 3RD ITEM DIFFERS | -.479 | -.620 | .761 |
| WEIGHT 3RD ITEM DIFFERS | .502 | .535 | .557 |
| D MINUS P | -.478 | 0. | .005 |

THERE ARE NØ SIGNIFICANT (.25) SPECIFICATIONS OF Q:
DØ YØU WANT TO SEE THE CØNDITIØNAL Q'S ('Y' ØR 'N')
?Y

I

| | | | |
|---|---|---|---|
| Q 3RD ITEM PØSITIVE | .004 | -.589 | .767 |
| WEIGHT 3RD ITEM PØS | .462 | .078 | .262 |
| Q 3RD ITEM NEGATIVE | -.055 | -.627 | .741 |
| WEIGHT 3RD ITEM NEG | .036 | .387 | .182 |
| CHI-SQUARE SPECIFICATIØN | .172 | .164 | .167 |
| SIGNIFICANCE, 1D.F. | .10 | .10 | .10 |

ARE YØU DØNE?
?Y

H. The user responds affirmatively to a DECØMPØSITIØN ØF Q. Partial and Differential Q's are produced in the subsequent table (along with the corresponding weights). Now look at Q for AGE vs READ--it disappears (0.000) when EDUC is controlled (EDUC tied) and becomes moderately negative or stronger for the differential (EDUC differs). Hajda's conclusion that education "EXPLAINS" the relationship between reading and age is justified.

I. A "specification" analysis is given for the raw data. There are no significant specifications. The analysis complete, the user says he is done. A novice who didn't understand all the logic in the analysis above--especially with respect to the terminology--now may want to use WORDS to help him write a report on the above analysis. A run of WORDS follows:

```
J   ?LØAD(WØRDS)PUBLIC
    ?XEQ
    WØULD YØU LIKE ME TØ WRITE THE FIRST DRAFT ØF YØUR REPØRT?
    ?YES

    PLEASE TYPE IN THE NAMES ØF YØUR INDEPENDENT, DEPENDENT,
    AND TEST VARIABLES.  (IF THEY ARE MØRE THAN 10 LETTERS LØNG
    USE ABBREVIATIØNS).

    WHAT IS THE NAME ØF YØUR DEPENDENT VARIABLE, Y?
    ?READ
K
    WHAT IS THE NAME ØF YØUR INDEPENDENT VARIABLE, X?
    ?AGE

    AND WHAT IS THE NAME ØF YØUR TEST VARIABLE, Z?
    ?EDUC

    DØ YØU WANT INPUT INSTRUCTIØNS?
    ?NØ
                        ENTER A,B,C,D,E,F,G,H
L   A
    ?215,263,373,453,335,54,133,24
    DID YØU ENTER YØUR DATA CØRRECTLY?
    ?YES

    HERE WE GØ!


          FIRST, LET'S LØØK AT THE ZERØ ØRDER RELATIØNSHIP
    BETWEEN AGE AND READ.
M
    WE SEE THAT THERE IS A LØW NEGATIVE ASSØCIATIØN BETWEEN
    AGE AND READ (Q= -0.24).

          NØW LET'S LØØK AT THE ZERØ ØRDER ASSØCIATIØNS
    INVØLVING EDUC.
    WE SEE THAT THERE IS A VERY STRØNG NEGATIVE ASSØCIATIØN BETWEEN
    EDUC AND AGE (Q= -0.62) AND A VERY STRØNG PØSITIVE ASSØCIATIØN
    BETWEEN EDUC AND READ (Q= 0.76).
```

J. The user loads and asks for the execution of the program, WORDS. He answers that he wants help in writing his report.

K. As in 2X2X2, he enters the dependent, independent and test (control) variables, this time not naming his dichotomy categories since this program merely treats each dichotomy as HIGH vs LOW.

L. Again refusing user instructions (he should ask for them the first time he runs) he enters the same data as he did for 2X2X2.

M. The zero order associations are given in a verbal report form. Evidence is quickly had as to the possibility of "explanation" of the zero order (AGE,READ) relationship due to the opposite signs of the other zero order relationships between AGE and EDUC and between READ and EDUC.

```
N  FRØM WHAT WE KNØW ALREADY, WHAT DØ YØU THINK WILL
   HAPPEN IØ THE ASSØCIATIØN BETWEEN AGE AND
   READ WHEN WE CØNTRØL FØR EDUC?  TYPE IN THE
   NUMBER NEXT TØ YØUR PREDICTIØN.

   IT WILL BECØME LESS NEGATIVE      7
   IT WILL BECØME MØRE NEGATIVE      8
   IT WØN'T BE AFFECTED MUCH         9

   WHAT DØ YØU THINK?
   Z
   ?7
   THAT'S RIGHT
   SINCE THE ASSØCIATIØNS HAVE ØPPØSITE SIGNS, THE
   NEGATIVE ASSØCIATIØN BETWEEN AGE AND
   READ WILL PRØBABLY BECØME LESS NEGATIVE WHEN
   EDUC IS CØNTRØLLED.

O  ENØUGH SPECULATIØN.  LET'S LØØK AT THE PARTIAL.

   WELL, WHAT DØ YØU KNØW!  THE PARTIAL Q ASSØCIATIØN BETWEEN
   AGE AND READ CØNTRØLLING FØR EDUC IS
     0.00, NØ ASSØCIATIØN.
   THUS, EDUC EXPLAINS THE RELATIØNSHIP BETWEEN
   AGE AND READ.  THAT IS, WHILE AGE AND
   READ HAVE A LØW NEGATIVE ASSØCIATIØN, IT BECØMES
   NEGLIGIBLE WHEN EDUC IS CØNTRØLLED.

P  BUT WE AREN'T THRØUGH YET.  THAT PARTIAL RELATIØNSHIP IS A
   SØRT ØF AVERAGE ØF THE AGE AND READ RELATIØNSHIPS
   IN THE TWØ CATEGØRIES ØF THE TEST VARIABLE, EDUC.
   THEY MIGHT BE RADICALLY DIFFERENT, BUT WE HAVE NØ WAY ØF
   PREDICTING FRØM THE PATTERN ØF ZERØ ØRDER ASSØCIATIØNS.

   HERE THEY ARE:  AMØNG THØSE SCØRED PLUS ØN EDUC
   THERE IS NØ ASSØCIATIØN BETWEEN AGE
   AND READ (Q=  0.00).  AMØNG THØSE SCØRED MINUS
   ØN EDUC THERE IS A NEGLIGIBLE NEGATIVE ASSØCIATIØN BETWEEN THEM.
   (Q=  -0.06).

   BY MAKING A SIGNIFICANCE TEST FØR THE DIFFERENCE IN THE
   TWØ VALUES OF Q, WE CAN TEST WHETHER THE DIFFERENCE BETWEEN
   THE TWØ ASSØCIATIØNS IS RELIABLE.  THE DIFFERENCE FAILS TØ
   REACH THE 0.025 SIGNIFICANCE LEVEL.  THUS, WE HAVE NØ FIRM
   EVIDENCE THAT EDUC SPECIFIES THE RELATIØNSHIP BETWEEN AGE
   AND READ AND HAVE ABØUT THE SAME ASSØCIATIØN IN EACH LEVEL.
```

N. The computer checks to see if the user understands this early indication of the effect of the control by the test variable. He does!

O. Then the computer analyses by the "partial Q" and comes up with a concrete report result--the EXPLANATION.

P. WORDS then checks the possibility of SPECIFICATION by calculating the positive and negative associations. A significance test, chi-square, is reported as not significant at the 0.025 level.

Q { THAT'S IT! ALL YØU NEED TØ DØ NØW IS TØ SCRATCH UP SØME BIBLIØGRAPHIC REFERENCES. HØWEVER, I MADE A FEW ARBITRARY DECISIØNS IN MY ANALYSIS, AND YØU ØUGHT TØ LØØK AT THE DATA ØN YØUR ØWN.
WHY DØN'T YØU TRY 2X2X2?

Q. A final pedogogical note!

## B. A "Specification" Variable

A second example of the use of 2X2X2 and WORDS for undergraduate research on published data is briefly presented to illustrate the flexibility of the programs. The Stauffer data from a 1954 national survey (Samuel A. Stauffer, "Communism, Conformity, and Civil Liberties," Doubleday, New York, 1955) indicated a subcultural difference in Southern voting behavior from the general behavior in the U.S. This difference is reflected in the preference for Eisenhower as compared to party preference. The new data being available, along with dichotomies, one can jump into 2X2X2 and WORDS.

A

```
SERVICE?
CPS
?xxx,yyy,zzz,name
GØØD AFTERNØØN;  USER 09;  TIME 16:30:50  10/28/71;
?LØAD(2X2X2)PUBLIC
?XEQ
```

B

```
DØ YØU WANT USER'S INSTRUCTIØNS?  'Y' ØR 'N'
?N
SUPPLY LABELS ØR ACCEPT X,Y,&T ('LABELS' ØR 'XYT')
?LABELS

NØW YØUR LABELS
FØR DEPENDENT VARIABLE, Y:
?PREF,IKE,STEV

FØR INDEPENDENT VARIABLE,X:
?REGIØN,SØUTH,ØTHER

FØR TEST VARIABLE, T:
?PARTY,REP,DEM
DID YØU ENTER YØUR VARIABLES CØRRECTLY?
TYPE 'Y' IF YØU DID,  'N' IF YØU WANT TØ RE-DØ THEM
?Y

ENTER A,B,C,D,E,F,G,H,
A
?26,231,26,1061,505,327,930,297
```

RAW DATA

| PARTY | REGIØN | /PREF/ STEV | IKE | |
|---|---|---|---|---|
| REP | SØUTH | 26 | 231 | 257 |
| | ØTHER | 26 | 1061 | 1087 |
| | | 52 | 1292 | |
| DEM | SØUTH | 505 | 327 | 832 |
| | ØTHER | 930 | 297 | 1227 |
| | | 1435 | 624 | |
| | | | | 3403 |

```
IS THIS TABLE CØRRECT?  ('Y' ØR 'N')
?Y
```

A. The user logs-in, calls 2X2X2, and asks to run the program.

B. The labels and data are fed in and 2X2X2 echoes out the contingency table. A request for an editorial check brings a response that all is O.K.--the dependent variable being PREF (voting preference) with the dichotomy categories of IKE and STEV (for Eisenhower and Stevenson); the independent variable REGION with dichotomy categories of SØUTH and ØTHER; and the test variable PARTY with dichotomy categories of REP and DEM.

C

```
DØ YØU WANT THE MARGINAL PERCENTAGES?
?Y

                    MARGINAL PERCENTAGES

                                REGIØN          PREF          PARTY
                               PERCENT        PERCENT       PERCENT
UPPER LIMIT (.95)               33.6           58.0          41.1
SAMPLE VALUE                    32.0           56.3          39.5
LØWER LIMIT (.95)               30.4           54.6          37.9
```

D

```
DØ YØU WANT A PERCENTAGE TABLE?
?Y

                        PERCENT IKE
                          REGIØN
                     ØTHER          SØUTH
          REP        97.6           89.9
                     (1087)         (257)
PARTY
          DEM        24.2           39.3
                     (1227)         (832)
```

E

```
DØ YØU WANT ZERØ ØRDER Q'S
?Y

                ZERØ ØRDER Q LIMITS AND SIGNIFICANCE
                      Q(REGIØN,PREF)  Q(PARTY,REGIØN)  Q(PARTY,PREF)
UPPER LIMIT (.95)        -.078           -.420            .975
SAMPLE VALUE             -.149           -.482            .966
LØWER LIMIT (.95)        -.219           -.544            .956
```

F

```
DØ YØU WANT THE DECØMPØSITIØN ØF Q?
?Y

                    DECØMPØSITIØN ØF Q

                                PAIR
                      (REGIØN,PREF)  (REGIØN,PARTY)  (PARTY,PREF)
=============================================================
Q ZERØ ØRDER             -.149          -.482           .966
Q 3RD ITEM TIED           .272          -.589           .971
WEIGHT 3RD ITEM TIED      .389           .371           .593
Q 3RD ITEM DIFFERS       -.417          -.419           .958
WEIGHT 3RD ITEM DIFFERS   .611           .629           .407
D MINUS P                -.688           .171          -.012
-------------------------------------------------------------
```

C. The MARGINAL PERCENTAGES all fall within the 30-70 percent range making analysis feasible.

D. The requested PERCENTAGES indicate different trends in SØUTH and ØTHER as well as in the Republican and Democratic party with respect to voting preference for IKE.

E. Zero order Q's show a low negative association between REGIØN and PREF, a moderate negative association between PARTY and REGIØN, and a very strong positive association between PARTY and PREF (the latter being not too shocking). A decomposition of the Q's seems called for.

F. The association between REGIØN and PREF turns moderately positive when controlling with tied pairs on the test variable PARTY (partial Q) while the differential Q (differing pairs for PARTY) showed moderately negative. This result (different signs--large difference) begs looking at the specification possibility. What happens in each PARTY?

G

```
YØU HAVE A SIGNIFICANI SPECIFICATIØN ØF Q:  HERE IS THE STØRY
Q3RD ITEM PØSITIVE          -.641          -.699          .864
WEIGHT 3RD ITEM PØS          .027           .341          .066
Q 3RD ITEM NEGATIVE          .339           .296          .984
WEIGHT 3RD ITEM NEG          .362           .031          .527
CHI-SQUARE SPECIFICATIØN 107.789         53.230        18.626
SIGNIFICANCE, 1 D.F.         .001           .001          .001
```

H

```
?LØAD(WØRDS)PUBLIC
?XEQ

WØULD YØU LIKE ME TØ WRITE THE FIRST DRAFT ØF YØUR REPØRT?
?YES

PLEASE TYPE IN THE NAMES ØF YØUR INDEPENDENT, DEPENDENI,
AND TEST VARIABLES.  (IF THEY ARE MØRE THAN 10 LETTERS LØNG
USE ABBREVIATIØNS).

WHAT IS THE NAME ØF YØUR DEPENDENT VARIABLE, Y?
?PREF

WHAT IS THE NAME ØF YØUR INDEPENDENT VARIABLE, X?
?REGIØN

AND WHAT IS THE NAME ØF YØUR TEST VARIABLE, T?
?PARTY
DØ YØU WANT INPUT INSIRUCTIØNS?
?NØ
                ENTER A,B,C,D,E,F,G,H
A
?26,231,26,1061,505,327,930,297
DID YØU ENTER YØUR DATA CØRRECTLY
?YES
```

I

```
HERE WE GØ!

      FIRST, LET'S LØØK AT THE ZERØ ØRDER RELATIØNSHIP
BETWEEN REGIØN AND PREF.

WE SEE THAT THERE IS A LØW NEGATIVE ASSØCIATIØN BETWEEN
REGIØN AND PREF (Q= -0.15).

      NØW LET'S LØØK AT THE ZERØ ØRDER ASSØCIATIØNS
INVØLVING PARTY.
WE SEE THAT THERE IS A MØDERATE NEGATIVE ASSØCIATIØN BETWEEN
PARTY AND REGIØN (Q=  -0.48) AND A VERY STRØNG PØSITIVE ASSØCIATIØN
BETWEEN PARTY AND PREF (Q= 0.97).
```

G. The specification is significant. Q, for the Republican party only, between REGIØN and PREF is strongly negative while Q, for the Democratic party, between REGIØN and PREF is moderately positive--an even bigger association difference than the partial and differential Q's showed. There was a subcultural difference between the South and the rest of the nation!

This is difficult for the novice to put into a report, so WORDS is called.

H. WORDS is called, the program is requested to run, and the variable names and raw data are entered correctly.

I. Zero order relationships are calculated and one report-like conclusion is reached: "There is low negative association between REGIØN and PREF (Q = -0.15)." The analysis now turns to see what effect PARTY has on this association.

FRØM WHAT WE KNØW ALREADY, WHAT DØ YØU THINK WILL
HAPPEN TØ THE ASSØCIATIØN BETWEEN REGIØN AND
PREF WHEN WE CØNTRØL FØR PARTY? TYPE IN THE
NUMBER NEXT TØ YØUR PREDICTIØN.

IT WILL BECØME LESS NEGATIVE 7
IT WILL BECØME MØRE NEGATIVE 8
IT WØN'T BE AFFECTED MUCH 9

J

WHAT DØ YØU THINK
Z
?7
THAT'S RIGHT
SINCE THE ASSØCIATIØNS HAVE ØPPØSITE SIGNS, THE
NEGATIVE ASSØCIATIØN BETWEEN REGIØN AND
PREF WILL PRØBABLY BECØME LESS NEGATIVE WHEN
PARTY IS CØNIRØLLED.

ENØUGH SPECULATIØN. LET'S LØØK AT THE PARTIAL.

WELL, WHAT DØ YØU KNØW? THE PARTIAL Q ASSØCIATIØN BETWEEN
REGIØN AND PREF CØNTROLLING FØR PARTY IS
0.27, A LØW PØSITIVE ASSØCIATIØN.

K

THE INTRØDUCTIØN ØF PARTY INTØ THE ANALYSIS
HAS LEAD US TØ CØMPLETELY REVISE ØUR THINKING ABØUT
REGIØN AND PREF. AT THE ZERØ ØRDER LEVEL, WE
SAW A LØW NEGATIVE ASSØCIATIØN BETWEEN THEM, BUT
WHEN PARTY IS CØNIRØLLED, A LØW PØSITIVE
ASSØCIATIØN APPEARS.

BUT WE AREN'T THRØUGH YET. THAT PARTIAL RELATIØNSHIP IS A
SØRT ØF AVERAGE ØF THE REGIØN AND PREF RELATIØNSHIPS
IN THE TWØ CATEGØRIES ØF THE TEST VARIABLE, PARTY.
THEY MIGHT BE RADICALLY DIFFERENT, BUT WE HAVE NØ WAY ØF
PREDICTING FROM THE PATTERN ØF ZERØ ØRDER ASSØCIATIØNS.

HERE THEY ARE: AMØNG THØSE SCØRED PLUS ØN PARTY
THERE IS A VERY STRØNG NEGATIVE ASSØCIATIØN BETWEEN REGIØN
AND PREF (Q= -0.64). AMØNG THØSE SCØRED MINUS

L

ØN PARTY THERE IS A MØDERATE PØSITIVE ASSØCIATIØN BETWEEN THEM
(Q= 0.34).

BY MAKING A SIGNIFICANCE TEST FØR THE DIFFERENCE IN THE
TWØ VALUES ØF Q, WE CAN TEST WHETHER THE DIFFERENCE BETWEEN
THE TWØ ASSØCIATIØNS IS RELIABLE.

THE DIFFERENCE IS SIGNIFICANT AT THE .025 LEVEL.
THUS, ØUR SECØND MAJØR CØNCLUSIØN IS THAT PARTY
SPECIFIES THE RELATIØNSHIP BETWEEN REGIØN AND
PREF. THAT IS, THE STRENGTH OF THE RELATIØNSHIP
BETWEEN REGIØN AND PREF VARIES WITH THE
LEVEL ØF PARTY.

J. The user is asked whether the other zero order Q's indicate what controlling by PARTY might mean to the (REGIØN, PREF) association. He correctly indicates that the signs of these associations (-.482 and .966) are opposite and should make the (REGIØN,PREF) Q less negative.

K. Analysis goes to the partial and differential Q's of (REGIØN,PREF) controlling for PARTY. A low positive association (less negative as predicted) of 0.27 is discovered. PARTY sure changed things.

L. The test of significance of a "specification" of the association by PARTY is calculated. A second major conclusion is reached and a report statement is generated: "The difference is significant---. That is, the strength of the relationship between REGIØN and PREF varies with --- PARTY." A subcultural difference is quantitatively expressed.

M
```
THAT'S IT!  ALL YØU NEED TØ DØ NØW IS TØ SCRATCH UP
SØME BIBLIØGRAPHIC REFERENCES.  HØWEVER, I MADE A FEW ARBITRARY
DECISIØNS IN MY ANALYSIS, AND YØU ØUGHT TØ LØØK AT THE DATA ØN
YØUR ØWN.
WHY DØN'T YØU TRY 2X2X2?
?LØGØUT
TIME 16:47:47;   TIME USED:   CPU  00:00:02;   TERM 00:16:56;   PAGE 00:46:33
```

M. The usual pedagogic departure.

A better understanding of the logic of this approach can be found in James A. Davis, "Elementary Survey Analysis," Prentice-Hall, Englewood Cliffs, N.J., 1971. The data for the above two examples were taken from a preliminary draft of this book.